

Shall I Encode Thee In DNA? Sonnets Stored On Double Helix

by ADAM COLE

January 24, 2013 3:37 AM

Audio for this story from Morning Edition will be available at approximately 9:00 a.m. ET.



Attributed to John Taylor/National Portrait Gallery

English critic Samuel Johnson once said of William Shakespeare "that his drama is the mirror of life." Now the Bard's words have been translated into life's most basic language. British scientists have stored all 154 of Shakespeare's sonnets on tiny stretches of DNA.

It all started with two men in a pub. Ewan Birney and Nick Goldman, both scientists from the European Bioinformatics Institute, were drinking beer and discussing a problem.

Their institute manages a huge database of genetic information: thousands and thousands of genes from humans and corn and pufferfish. That data — and all the hard drives and the electricity used to power them — is getting pretty expensive.

"The data we're being asked to be guardians of is growing exponentially," Goldman says. "But our budgets are not growing exponentially."

It's a problem faced by many large companies with expanding archives. Luckily, the solution was right in front of the researchers — they worked with it every day.

"We realized that DNA itself is a really efficient way of storing information," Goldman says.

DNA is nature's hard drive, a permanent record of genetic information written in a chemical language. There are just four letters in DNA's alphabet — the four nucleotides commonly abbreviated as A, C, G and T.

When these letters are arranged in different ways, they spell out different instructions for our cells. Some 3 billion of those letters make up the human genome — the entire instruction manual for our existence. And all that information is stuffed into each cell in our bodies. DNA is millions of times more compact than the hard drive in your computer.

The challenge before Goldman and his colleagues was to make DNA store a digital file instead of genetic information.

"So over a second beer, we started to write on napkins and sketch out some details of how that might be made to work," Goldman says.

They started with a text file of one of Shakespeare's sonnets. In the computer's most basic language, it existed as a series of zeroes and ones. With a simple cipher, the scientists translated these zeroes and ones into the letters of DNA.

And then they did the same for the rest of Shakespeare's sonnets, an audio clip of Martin Luther King Jr.'s "I Have a Dream" speech, and a picture of their office. They sent that code off to Agilent Technologies, a biotech company. Agilent synthesized the DNA and mailed it back to Goldman.

"My first reaction was that they hadn't done it properly, because they sent me these little tiny test tubes that were quite clearly empty," Goldman says.

But the DNA was there — tiny specks at the bottom of the tubes. To

read the sonnets, they simply sequenced the DNA and ran their cipher backward. All the files were 100 percent intact and accurate.

They published their results in the journal *Nature*, joining other groups who have experimented with DNA storage. George Church, a geneticist at Harvard who helped start the Human Genome Project, encoded an HTML file of his latest book into DNA earlier this year.

Goldman and Birney's method included greater redundancies and overlapping stretches of DNA to prevent against errors. They say the process would be easy to scale up.

If you took everything human beings have ever written — an estimated 50 billion megabytes of text — and stored it in DNA, that DNA would still weigh less than a granola bar.

"There's no problem with holding a lot of information in DNA," Goldman says. "The problem is paying for doing that."

Agilent waived the cost of DNA synthesis for this project, but the researchers estimate it would normally cost about \$12,400 per megabyte.

"It's an unthinkable large amount of money ... at the moment," Goldman says.

Goldman and other scientists who are dabbling in DNA storage know that DNA synthesis costs are dropping rapidly. In a decade or so, they say it may be more cost effective for large companies to keep a DNA archive than to maintain and update a roomful of hard drives.